Death and Lightness: Using a Demographic Model to Find Support Verbs

Mark Dras & Mike Johnson
Microsoft Research Institute & Department of Computing
Macquarie University
North Ryde NSW 2109
Australia
{markd,mike}@mpce.mq.edu.au

Abstract

Some verbs have a particular kind of binary ambiguity: they can carry their normal, full meaning, or they can be merely acting as a prop for the nominal object. It has been suggested that there is a detectable pattern in the relationship between a verb acting as a prop (a SUPPORT VERB) and the noun it supports.

The task this paper undertakes is to develop a model which identifies the support verb for a particular noun, and by extension, when nouns are enumerated, a model which disambiguates a verb with respect to its support status. The paper sets up a basic model as a standard for comparison; it then proposes a more complex model, and gives some results to support the model's validity, comparing it with other similar approaches.

1 Introduction

It is well-known that some verbs have a binary ambiguity: consider the sentences *Kim took a photograph of Dale* and *Kim took a painting of Dale*. The former can be a paraphrase of *Kim photographed Dale* while the latter has no such paraphrase. This is because, in one reading of the first example, the lexeme *take* is only acting as a prop or support for a content-bearing noun, a capacity first noted by Jespersen (1942); while in the second example the verb has only its full meaning of *gain possession*. It has been noted that many support verbs (SVs), like *take* and *make*—as in *make a distinction* (equivalent to *distinguish*)—are quite productive, being able to act as support verbs for a number of different nouns; and investigations have suggested that there may be a pattern in the relationship between SV and noun (e.g. Makkai, 1977; Wierzbicka, 1982). Discovering the correspond! ence between SV and noun can help disambiguate the verb, deciding whether it is acting as an SV or not; that is the aim of this paper.

The concept of ambiguity here is similar to that of word-level sense ambiguity (see, for example, Yarowsky, 1992), rather than to higher-level ambiguity such as that of garden-path sentences (Gibson, 1995). Because an SV by itself does not represent a separate concept—the whole construction *take a walk* represents a single concept of walking—while a full verb can, removing this type of ambiguity from input is particularly important for determining mappings in areas where input text is translated into another form, such as to text in another language, in machine translation (Danlos and Samvelian, 1992), or to a meaning representation (Meteer, 1991); it is also useful for dealing with multiword constructions in language, like idioms (Abeille, 1988; Storrer and Schwall, 1993).

More generally, identifying a verb as an SV indicates its lack of propositional content, and so can contribute to more accurate readability measures, such as lexical density (described in Halliday, 1985). Knowing whether a word lacks content is similarly important in the area of information retrieval, in explicitly constructing stoplists (Salton, 1988), on the assumption that content-free words should be deleted from the search space of key terms. These lists can be made more comprehensive by recognising the similar lack of content in SVs and the closed-class words, which are traditionally considered to comprise the set of content-free words (Halliday, 1985). Another area of potential use is in style checking, where it is generally recommended that SVs be removed for reasons of clarity (Kane, 1983); for example, [1a] becomes [1b].

1a. It is important for teachers to have a knowledge of their students.

1b. It is important for teachers to know their students.

Characterisation and identification techniques for SVs have used both purely semantic methods and more syntactic, surface-based ones. An example of the former is given in Wierzbicka's 1982 paper, where a set of semantic rules is proposed to de-

termine the SV that corresponds to a particular noun, concentrating on explanations of phenomena like why one can *have a drink* but not **have an eat*. Her analysis of these phenomena leads to rules like:

2. The support verb is *have* if the nominalisation represents an action aiming at a perception which could cause one to know something and which would not cause one to feel bad if it didn't.

The surface-based approaches aim to overcome the laborious nature of determining such semantic rules by assuming that the syntactic structure reflects enough of the semantics to make a surface statistical analysis possible. Fontenelle (1993) proposed a surface-based approach, which uses the work of Smadja (1991) on collocation relations in text. His method, however, requires multi-lingual machine readable dictionaries, which may not be readily available, and the prior division of words into sets according to Mel'cuk's Meaning-Text Theory (c.f. Mel'cuk and Zholkovsky, 1988; Steele, 1990).

A more recent example of the surface-based approach is the statistical technique proposed by Grefenstette and Teufel (1995). A statistical analysis sounds intuitively plausible, given that it has been suggested (Halliday, 1985) that there is a relationship between frequency of a word, a surface phenomenon, and its content-freeness. However, Grefenstette and Teufel's (1995) statistical technique only uses frequency with respect to a particular noun (which I call LOCAL INFORMATION), rather than any more general notion of frequency. So, for example, in identifying an SV for *demand*, Grefenstette and Teufel look only at the frequency of co-occurrence of various verbs with *demand*. As a result, in their system *meet* is chosen as the corresponding SV as it is the most frequently co-occurring verb; *make*, the actual SV, is ranked lower. However, knowing that *make* is a generally productive SV would lead it to be a more obvious candidate, despite its lower ranking. A probabilistic argument to this effect is given in section 3; the key approximation on which it relies, the use of data with respect to all other SV constructions (LOCAL INFORMATION), is drawn from a model in demographic statistics, and is outlined first in section 2.

2 Mortality

This section looks at a standardisation model used in mortality which is a useful one for SVs: it provides a way of combining information about a subpopulation—the TARGET POPULATION—with a larger population which provides more information—the STANDARD POPULATION. An overview of the theory is given below; more detail can be found in standard demographic texts such as Pollard *et al* (1981). The method described in this paper combines local and global information about SVs in a similar way; the correspondence will be discussed in more detail in Section 3.

2.1 Standard populations

The most easily obtainable mortality rate, the crude death rate (CDR), is calculated by dividing the total number of deaths for a population by the total size of the population. However, this does not accurately reflect the mortality experienced by the population: Pollard *et al* (1981) discuss the situation of Maori and non-Maori populations of New Zealand in 1966, where the Maori population had a lower CDR despite having higher mortality rates for every age group.

The explanation for this discrepancy comes from the different profiles of each population: the age categories which experience the lowest rates have higher population sizes, weighting the overall population rate so that it also is lower. So the Maori population has a much higher number of young people, who have lower death rates; this produces a lower overall rate, as the CDR effectively weights the measurement by the distribution of the Maori population. Using a common (or standard) population is one way of removing this bias.

2.2 Indirect Standardised Death Rate (ISDR)

One standard demographic technique for producing a figure comparable between populations is to apply age-specific mortality rates from the standard population to the corresponding age brackets of the target population, giving the number of deaths that would be expected in the target population if the levels of mortality in the standard population were being experienced. These expected deaths are summed, and used in the calculation of the standardised mortality ratio, which is equal to actual deaths for the target population divided by expected deaths; it represents the degree above or below expectation to which deaths actually occurred. This ISDR is often the preferred measure of standardisation when the target population is too small to accurately calculate age-specific mortality rates, using as it does those of the standard population in their place.

There is no one definitive standard population for two given target populations. One frequently chosen standard population is the union of the two target populations: for example, when comparing Maori mortality with non-Maori mortality in New Zealand, the total New Zealand population was used as standard.

3 A Probabilistic Model

This section describes a probabilistic model for the prediction of support verbs, along with the approximations and assumptions being made; these are then justified by recourse to the demographic model described in section 2.

The most likely support verb for a given nominalisation is defined as that verb which has the highest probability of being an SV for that nominalisation; taking the point estimate of this probability, the most likely support verb for a nominalisation is that verb which has the highest frequency of occurrence as an SV with the nominalisation. That is:

$$\mathrm{SV}(j) \quad = \quad argmax_{i \in V} f_{ij} \tag{1}$$

where

- $SV(j)$ = most likely SV for nominalisation j
- f_{ij} = frequency with which verb i appears to be supporting nominalisation j

This quantity f_{ij} is, of course, unknown, as there are no corpora tagged for verb lightness—that tagging is the purpose of the identification method proposed in this paper and others. Now, f_{ij} can be rewritten as

$$\mathrm{f}_{ij} \quad = \quad \mathrm{m}_{ij} p_{ij} \tag{2}$$

where

- m_{ij} = number of occurrences of verb i governing[1] nominalisation j
- p_{ij} = Pr (verb i is acting as an SV | nominalisation j)

3.1 Basic model

Again, p_{ij} is unknown and cannot be estimated directly. One approach is to make the admittedly inaccurate assumption that p_{ij} is equal to 1 for all i and j. That is, the verb chosen to be the SV is simply that one which most frequently governs the verb in the chosen training corpus. Then

$$\mathrm{f}'_{ij} \quad = \quad \mathrm{m}_{ij} \tag{3}$$

which gives

$$\mathrm{SV}(j) \quad = \quad argmax_{i \in V} m_{ij} \tag{4}$$

[1] I use 'govern' in the sense that if X is a complement of Y, then Y governs X; this is in line with Mel'cuk (1988).

Grefenstette and Teufel (1995) effectively use this assumption, with the additional modification of restricting the count m_{ij}, by only counting those occurrences where the SV construction has similar characteristics to the equivalent full verb. So, for example, the preposition qualifying the noun in the SV construction is generally the same as the preposition attached to the full verb (*make a decision to ..., decide to ...*); they use this type of information, when collecting data, to give a more accurate m_{ij}.

In this paper I will only be looking at the gain that can be made from attempting to estimate p_{ij}, so I will be using this definition of $SV'(j)$ as the main basis for comparison. I will, however, also compare the results against the model of Grefenstette and Teufel (1995), to compare the degree of improvement expected of each over the basic model.

3.2 Global Information Model

Now, an approximation for p_{ij} suggested by the demographic model above is to use the unconditional probability over all nominalisations—call this p_i. So:

$$\mathrm{f}''_{ij} \quad = \quad \mathrm{m}_{ij} p_i \tag{5}$$

where

$$\begin{aligned} \mathrm{p}_i &= Pr(verb.i.is.acting.as.an.SV) \\ &= \sum_j m_{ij} p_{ij} / \sum_j m_{ij} \end{aligned} \tag{6}$$

The case for using such a significant approximation here is the same as the case for using it in the demographic model described in section 2; it is the approximation around which the model is built. The use of unconditional probabilities as an approximation to the conditional ones parallels, in the demographic model, the use of the standard or global population rates when calculating statistics on the sub- or target population. The correspondence between elements of the two models is given in Table 1.

The reason behind using the global rates in both cases is similar—the local probabilities cannot accurately be estimated. A difference, however, can be noted. In the demographic model, the local and global probabilities of dying will both usually follow typical mortality curves: high rates at birth, declining until late teens, an 'accident hump' of higher mortality, another decline, and then increasing with middle and old age. But, in the language model, the conditional and unconditional probabilities are less innately similar. The conditional probabilities will generally be more dichotomous: for a given nominalisation, a verb will either (virtually) always or

Mortality		Lightness	
description	instance	description	instance
target population	Maori population	local information (data for given nominalisation, such as *make*)	all verbs governing given nominalisation, such as *make*
standard population	NZ population	global information (data for all nominalisations)	all verbs governing all nominalisations
age category	ages 15-25	verb	instances of *make* as a governing verb
target population mortality rate		conditional probability p_{ij}	
standard population mortality rate		unconditional probability p_i	

Table 1: Correspondence between mortality and lightness models

(virtually) never be an SV. The global, unconditional probabilities will, on the other hand, be a more mixed distribution: for example, *make* may have an unconditional probability of being an SV of 0.3, *have*, a probability of 0.23, and so on.

It should also be noted, however, the mortality curves of the demographic model can actually be very dissimilar also—the mortality rate distribution for the target population may lack an accident hump, or the probability of dying may approach 1 much faster and earlier than in the global mortality distribution, producing an effect similar to that in the language situation described. The approximation technique is fairly robust, to allow for this, and is not greatly affected by the choice of the global population or rates (see Pollard, 1981: 72). In any case, it is more accurate than the assumption of p_{ij} equaling 1: the p_i are a ranking of the likelihood of a verb being an SV when no context is known, meaning that more likely candidates can be identified, whereas the basic model gives no such indication.

Estimating these unconditional probabilities relies on an assumption that support verbs are productive to some extent, an assumption which appears to hold true for at least the major support verbs—for example, *make* acts as an SV for *attempt*, *criticism*, *decision*, *error*, *judgment*, and many other nominalisations—with a corollary to the assumption, that other non-support verbs will not exhibit the same generality across nominalisations, and this seems to be borne out by inspection of the global information described in section 4.

Then, given this assumption, the unconditional probabilities can be estimated by, for the purposes of this estimation only, treating all occurrences of verbs governing nominalisations in the corpus as acting as support verbs, and aggregating these to give the unconditional probability. That is:

$$\mathrm{p}'_i = \sum_j m_{ij} / \sum_i \sum_j m_{ij} \qquad (7)$$

This can be thought of as producing a global ordering of verbs, in which the rankings approximate the likelihood of acting as an SV because of the productivity of support verbs, which will leading to p_i and p'_i correlating reasonably well. Productive support verbs will tend to govern a range of different nominalisations and be ranked high in this ordering; their higher values of p'_i correspond to their higher likelihood of being support verbs as measured by p_i. A less productive SV, such as *bear*, will still have a low probability estimate p'_i. This approximation is not expected to give accurate estimates for p_i; it is only important that it correlate with p_i, as the process of choosing the most likely SV only requires that the ranking of verbs in order of the probability of being an SV be accurate. Again, an inspection of the global information described in section 4 seems to bear this out.

Given these approximations, the most likely SV under this model is given by:

$$\begin{aligned} \mathrm{SV}''(j) &= argmax_{i \in V} m_{ij} p_i \\ &= argmax_{i \in V} m_{ij} \sum_j m_{ij} / \sum_i \sum_j m_{ij} \\ &= argmax_{i \in V} m_{ij} \sum_j m_{ij} \qquad (8) \end{aligned}$$

4 Experimental work

4.1 Deriving local and global information

To gather both local and global information, the 1992 version of Grolier's encyclopedia of approximately 8 million words was used, tagged by the part-of-speech tagger developed by Brill (1993). A heuristic for producing the local information about a tar-

get population involved searching the corpus for the nominalisation, determining the verb for which the nominalisation was the direct object, and tallying the relative frequency of these verbs.

Grefenstette and Teufel (1995) note that a confounding factor in the local information, when picking out nominalisations and their governing verbs, is that the nominal may have become CONCRETISED. Generally, nominalisations represent an abstract concept, being essentially events represented in noun form; but it is possible for the nominal to represent a physical embodiment of that concept. For example:

3a. ABSTRACT: He made his formal proposal to the full committee.

3b. CONCRETISED: He put the proposal in the drawer.

The abstract and concretised versions will tend to have different governing verbs. However, if the assumption about productivity in Section 3.2 is true, and the global information is a good approximation to the innate lightness of a verb, the correct SV will be favoured over those associated with the concretised forms.

4.2 Generating nominalisations for global information

To construct the global information, data for all nominalisations is needed. A large list of nominalisations was derived in a partially automated manner from Longman's Dictionary of Contemporary English (LDOCE) using both built-in information and a heuristic: since a nominal is an event represented in noun form, the procedure used here for deriving a list of them involved looking for nouns with associated STEM VERBS; e.g., *decide* is the stem verb of *decision*. Some verbs have this information encoded in their entries: for example, *adjust* lists *adjustment* as its nominalisation; there were 257 verbs in this category. For others, an automatic orthographic heuristic that matched nouns with verbs produced a set of candidates, which was manually filtered to produce 1414 more deverbal nominalisations.

A set of support verb constructions and their constituent nominalisations was drawn from a range of sources—see Table 2 and bibliography for references—and used as the test set for the experiment. [2] The list of nominals did not cover some of the nominals from the test, so the local information was generated from the training corpus for each of the missing test set nominals and aggregated into the global information.

[2] These sources have assumed that the propositional meanings of the SV construction and the full verb are equivalent. This may be disputed in a number of cases, but for the purposes of this paper, the equivalence of the two meanings will be taken as indicated by the relevant source.

4.3 The test set and results

A system to identify support verbs for nominalisations, based on the global information model, was implemented by tabulating the lemmatised forms of all the verbs for which these nominals were the direct object. Candidate support verbs were ranked in order of values of SV"(j), and the maximum of these values chosen.

The test set and results are summarised in Table 2; the table contains:

- the source text;
- the corresponding verb, which the source can be rewritten as;
- the reference from which the source text was taken;
- the system's first choice candidate for support verb C1 for the source text's constituent nominalisation (i.e. the verb category with the highest expected number of light verbs, SV"(j));
- the system's second choice C2; and
- the ratio of the expected number of light verbs for the first and second choices.

A second system was implemented based on the basic model of section 3.1; for comparison, Table 3 gives the results of this system.

4.4 Discussion

4.4.1 Analysis of results of the global information model

Of the 18 examples, 13 choices of support verb match the corresponding one from the source text. Of the five cases where the chosen SV did not match the source verb, one was actually valid: *harm* had *cause* as the proposed alternative. This is an equally plausible support verb, and in any case, *do* was the second choice by only a small margin. This is true for a number of cases: where there is an alternative support verb to the one used in the source text, the second alternative represents another plausible choice, and the frequency ratio margin is small (for example, for *change* and *resemblance*).

In three cases lack of data is a problem, resulting in the three N/A values in Table 2: there are no occurrences of *snooze* or *shove* as direct objects of verbs in Grolier's, most probably because they belong to a more informal register than that used in encyclopedias. Similarly, *have a drink* is an informal phrase that would not normally be found in an encyclopedia, as evidenced by there being only one occurrence of a governing verb for *drink*.

Source Text	Verb	Choice C1	Choice C2	Ratio (C1/C2)	Reference
make an attempt	attempt	make	include	9.36	Dras, Dale (1995)
make a change	change	make	produce	1.85	Dras, Dale (1995)
make a concession	concede	make	include	11.47	Dras, Dale (1995)
make a demand	demand	make	create	1.03	Gref., Teufel (1995)
make a distinction	distinguish	make	have	3.04	Meteer (1991)
have a drink (of)	drink	become	N/A	N/A	Wierzbicka (1982)
have an effect (on)	affect	have	produce	3.04	Dras, Dale (1995)
have a feeling	feel	have	produce	3.27	Harris (1957)
make a gift (of)	give	have	include	9.89	Harris (1957)
do harm (to)	harm	cause	do	1.26	Huddleston (1968)
make a judgment	judge	make	have	2.43	Dras, Dale (1995)
have a knowledge (of)	know	have	use	12.36	Kane (1983)
make progress	progress	make	allow	64.33	Harris (1957)
make a proposal	propose	make	include	1.10	Gref., Teufel (1995)
bear a resemblance (to)	resemble	bear	have	2.64	Huddleston (1968)
give a shove (to)	shove	N/A	N/A	N/A	Harris (1957)
have a snooze	snooze	N/A	N/A	N/A	Harris (1957)
make use (of)	use	make	have	6.55	Dras, Dale (1995)

Table 2: Support verb candidates chosen by the system

Source Text	Verb	SV'(j)
make an attempt	attempt	make
make a change	change	undergo
make a concession	concede	make
make a demand	demand	meet
make a distinction	distinguish	make
have a drink (of)	drink	become
have an effect (on)	affect	have
have a feeling	feel	express
make a gift (of)	give	have
do harm (to)	harm	cause
make a judgment	judge	make
have a knowledge (of)	know	have
make progress	progress	make
make a proposal	propose	reject
bear a resemblance (to)	resemble	bear
give a shove (to)	shove	N/A
have a snooze	snooze	N/A
make use (of)	use	make

Table 3: Support verb candidates chosen under the basic model

The system's worst performance was with the nominalisation *gift*. This appears to have occurred as *gift* is frequently concretised, as in *She has a great gift which has astounded her teachers* or *This deal includes a free gift!* However, only one of the 18 cases appears to be affected in this way.

4.4.2 Comparison

So, allowing alternative SVs, and disregarding the cases where the genres of the test data differed from the genre of the training (encyclopedia) data, the success rate is 14 of 15 (93%), using a 66Mb corpus. By comparison, Grefenstette and Teufel (1995) achieve plausible SVs for 7 of 10 cases (70%), using a 134Mb corpus; and using the basic model described in section 3 achieves plausible support verbs for 10 of 15 cases. The higher results achieved by the method proposed in this paper are statistically significant at the 10% and 5% levels respectively. Stronger results may be obtained given more test data—it is difficult to do better than an improvement of 5% significance with only 15 cases. Developing a larger set will require further work, as there is often disagreement about the validity of equivalence between SV constructions and full verbs. The data do suggest, however, that this is worthwhile, particularly when it is noted that the higher success rate was achieved with a smaller corpus.

In general, the method seems to cover well both productive and idiomatic SV constructions. For example, *have* is productive in light verb constructions, and the high global frequency will give a relatively high expected lightness rate; however, it does not eliminate the possibility of a low-frequency verb (like *bear*) being a support verb in cases where the SV construction is strongly idiomatic (as in *bear a resemblance*), where the low frequency in the standard population is counterbalanced by a high frequency in the target population.

5 Conclusion

In the process of calculating expected SVs, a number of significant assumptions were made:

- that concretised nominals would not have a significant impact compared with the effect of the standardisation;
- that in initially constructing the global information, all verbs can be taken as light; and
- that the productivity of SVs allows the construction of a reasonable standard population.

Notwithstanding these considerations, the experimental results demonstrate that approximating the conditional probabilities, by the unconditional probabilities derived from a large number of nominalisations, provides accurate choices for support verbs for individual nominalisations in the test set. The accuracy of the method appears to be superior to existing statistical methods which use only local information; and the method also involves only minimal development effort, unlike existing semantic methods.

It is apparent that what constitutes a valid light verb construction depends on the genre and register of the text. Given that the test set was taken from a wide range of sources, more accurate results for this test set could no doubt be obtained by using a corpus that was more representative of general English. Also, more accurate results might be gained after further iterations of this process: once the most likely support verb in a given local information is determined, the global information can be regenerated using these, rather than the assumption of universal lightness of verbs. Further work will look at implementing this iterative process, developing a larger set of test data for evaluation purposes, and extending this method to other light constructions—light verbs with adjectival complements, and light nouns with post-modifiers.

Also, in order to successfully carry out a process of disambiguation on a random text, the coverage of nominalisations and SVs needs to be greater; a key aspect of future work is expanding the set of data to achieve this better coverage.

References

Abeille, Anne. 1988. "Parsing French with Tree Adjoining Grammar: some linguistic accounts" *Proceedings of COLING 1988.*

Brill, Eric. 1993. *A Corpus-based Approach to Language Learning.* PhD Thesis, University of Pennsylvania, PA.

Danlos, Laurence and P. Samvelian. 1992. "Translation of the predicative element of a sentence: category switching, aspect and diathesis." *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, 21-34.

Dras, Mark and Robert Dale. 1995. *Style and Semantic Density.* Unpublished manuscript.

Fontenelle, Thierry. 1993. "Using a bilingual computerized dictionary to retrieve support verbs and combinatorial information." *Acta Linguistica Hungaraica*, 41(1-4), 109-121.

Gibson, Edward. 1995. *Memory Limitations and Sentence Processing Breakdown.* MIT Press. Cambridge, MA.

Grefenstette, Greg and Simone Teufel. 1995. "Corpus-based Method for Automatic Identification of Support Verbs for Nominalizations". *Proceedings of EACL'95.*

Halliday, Michael A. K. 1985. *Spoken and Written Language.* Oxford University Press, Oxford.

Harris, Zellig. 1957. "Co-occurrence and transformation in linguistic structure". *Language*, 33, 293-340.

Huddleston, Rodney. 1968. *Sentence and Clause in Scientific English.* Communication Research Centre, University College, London.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles, Vol VI.* Allen and Unwin, London.

Kane, Thomas S. 1983. *The Oxford Guide to Writing.* Oxford University Press, New York, NY.

Makkai, Adam. 1977. "The passing of the syntactic age: A first look at the ecology of the English verb *take*". In *Linguistics at the crossroads* (ed. Adam Makkai *et al*) 79-104. Liviana, Padova.

Mel'cuk, Igor. 1988. *Dependency Syntax: Theory and practice.* SUNY Press. Albany, NY.

Mel'cuk, Igor and A. Zholkovsky. 1988. The Explanatory Combinatorial Dictionary. In *Relational Models of the Lexicon* (ed. M. Evans) Cambridge University Press. Cambridge, UK.

Meteer, Marie W. 1991. "The Implications of Revisions for Natural Language Generation". In *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, Kluwer Academic Publishers. Norwell, MA.

Pollard, A. H., Farhat Yusuf and G. N. Pollard. 1981. *Demographic Techniques.* Pergamon Press (Australia) Ltd. Rushcutters Bay, Australia.

Salton, Gerard. 1988. *Automatic Text Processing: The transformation, analysis and retrieval of information by computer.* Addison-Wesley. Reading, MA.

Smadja, Frank. 1991. "Macrocoding the lexicon with co-occurrence knowledge." In *Lexical Acquisition: Using On-Line Resources to Build a Lexicon* (ed. U. Zernik) Erlbaum Associates. Hillsdale, NJ.

Steele, J. (ed.) 1990. Meaning-Text Theory: Linguistics, Lexicography, and Implications. University of Ottawa Press. Ottawa, Canada.

Storrer, Angelika and Ulrike Schwall. 1993. "Description and Acquisition of Multiword Lexemes." *Proceedings of the Third International EAMT Workshop*, 35-50.

Wierzbicka, Anna. 1982. "Why Can You *Have a Drink* When You Can't **Have an Eat*?" *Language*, 58(4), 753-799.

Yarowsky, David. 1992. "Word sense disambiguation using statistical models of Roget's categories trained on large corpora". *Proceedings of the 14th International Conference on Computational Linguistics*, 454-460.